DOCUMENT RESUME

ED 390 286 FL 023 488

AUTHOR Puhl, Carol A.; Swartz, J. J.
TITLE A Semi-Structured English Oral Proficiency Test for Certification of Teachers in South Africa.
PUB DATE Apr 95
NOTE 8p.; Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (29th, Long Beach, CA, March 26-April 1, 1995).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *English; *English (Second Language); Foreign Countries; Interpersonal Communication; *Language Proficiency; Official Languages; Persuasive Discourse; Speech Skills; Standardized Tests; *Teacher Certification; Test Construction; *Testing Programs; Test Use; *Verbal Tests
IDENTIFIERS *South Africa

ABSTRACT

The construction, administration, and scoring of the Teachers Oral Proficiency Interview for South Africa (TOPISA) are described, and issues in its use are discussed. The test was developed to help standardize English proficiency levels of teachers in the post-apartheid, multicultural society. The rationale for a standardized test in this context is examined first. The three phases of the test (informal social discourse, expository discourse, argumentative discourse) are then described briefly, and the scoring standards and interpretation, which range from "below standard for teaching in South Africa" and "able to function in English at a near-native level of competence," are explained. A discussion of the theoretical foundations of the test, and its origins in testing practice, follows. Finally, some areas in which further research are needed are outlined, including clearer description of skills tested, validity (face, response, concurrent, content, predictive, and construct), reliability, elements defining proficiency, potential for adaptation from a direct to a semi-direct measure (e.g., taped interviews for later scoring), and potential for use with the 10 other official languages of South Africa. A brief bibliography is included. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A SEMI-STRUCTURED ENGLISH ORAL PROFICIENCY TEST FOR CERTIFICATION OF TEACHERS IN SOUTH AFRICA

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Dr Carol A Puhl and Dr J J Swartz
Faculty of Education, University of Stellenbosch
Stellenbosch 7600, South Africa

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Carol A Puhl
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Paper presented at TEACHERS OF ENGLISH TO SPEAKERS OF OTHER LANGUAGES CONFERENCE, Long Beach, California, 1 April 1995.

ED 390 286

There is a great need in South Africa, a country of 40 million people, for a sound, feasible and credible assessment of English language proficiency of pre-service teachers. Their English competency is a major factor in their effectiveness as teachers in a newly-democratised, multicultural nonracial society. Although there are 11 official languages, the envisaged language policy for the new South Africa emphasizes the paramount role of English as the likely medium of instruction in schools in general. Given that about 90% of South Africans have a mother tongue other than English, the need is vast.

WHY A STANDARDISED TEST?

Post-apartheid education is unified under one department of national education, with the same standards and rewards for all teachers based on their qualifications, which in turn determine appointments to posts, permanency of tenure in post, and possibility of advancement and remuneration. The English qualification must thus be assessed with great care.

In addition to employment reasons, the teacher's quality of oral English is bound to have a constant impact on pupils, who need good modeling of language in use. In fairness to the teachers themselves, they should not be put into situations in which they lack the language skills needed for them to cope.

In South Africa, there is a need for standardization in the important task of teacher certification in English because about 40 tertiary institutions in South Africa assume this responsibility. It seems that each situation has set its own standards in this area, with varying results for the country as a whole. What is urgently required is a test with universal application which can be used by all certification bodies and which will yield valid and reliable results. Now that the country has become a true democracy, attention can be given to improving education, which is one of the top three priorities of the Reconstruction and Development Programme (RDP).

Requirements for such a measure are various. It must be appropriate for teachers from diverse language and cultural situations; appropriate for diverse subjects; communicative; credible (seen as fair); without bias due to ethnicity, race, or gender; focused on required English language competencies for teaching and professional functioning; affordable; time-feasible (not too long to take or to score); comprehensive

FL023488

 BEST COPY AVAILABLE

(oral, reading, and writing skills). It must apply to teachers in elementary and secondary schools.

While a written component of the English certification test is set, given, and scored with confidence, the oral component is still subject to a great amount of individual tester discretion. This paper focuses on the oral component.

THE TOPISA

The University of Stellenbosch English Education Department has developed a model Teachers Oral Proficiency Interview for South Africa (TOPISA) that meets the above requirements. It has been used successfully for three years. It is informed by both the British model of adaptability of content and by the American emphasis on standardisation (Alderson, Krahnke & Stansfield, 1987). First, the test will be described in detail, and then examined theoretically. Possibilities for future research and development will be indicated.

The TOPISA examines teachers' command of structure, lexis (vocabulary), subject-specific terminology, register (appropriateness of language choice), relevant didactic functions (social, expository, argumentative), and fluency. These dimensions are among those given as essential by Stansfield, Karl, and Kenyon (1990). It seems related to another assessment instrument for teachers, the TOPT, or Texas Oral Proficiency Test, developed to certify teachers in French, Spanish or bilingual education (Stansfield, 1993).

Language tests, specifically oral proficiency tests, can take several forms according to various dimensions, including directness. A test can be direct, semi-direct, or indirect (Clark, 1979). Although the TOPISA is direct, a question for future investigation is raised: Can the TOPISA be adapted to form a taped, semi-direct measure that can be administered over time and distance? This would address several issues: the growing need for a national standard, under one department of education; the need to assess greater numbers of prospective teachers; the need for greater test credibility, to be seen as fair to all, irrespective of one's mother tongue.

ADMINISTRATION OF THE TOPISA

The TOPISA is administered in a 20-minute interview which is composed of three phases of controlled interaction. The interaction is semi-structured rather than completely spontaneous and free-flowing, while still allowing testees to talk in a natural manner. Administration of the TOPISA is done by two Interviewers/evaluators (Is) to two students, unknown to each other and from different subject backgrounds.

Phase I: Informal social discourse. Participants exchange appropriate greetings, introductions, and brief biographical information. They should return the greetings satisfactorily and introduce themselves to each other and to the Interviewers, telling briefly about where they are from, their major subject, leisure interests, goals, etc. All in the group may ask brief follow-up questions, as the setting is a relaxed chat.

The social discourse section is not trivial; it carries initial or opening messages of relationship, as opposed to content. These messages are crucial in determining classroom climate, which includes an atmosphere of encouragement and motivation. Such functions form part of daily classroom routine. They are also part of a teacher's role in communicating meaningfully with other teachers, in attending professional development conferences and seminars, and in dealing with parents and other community-based stakeholders in the educational process.

Phase II: Expository discourse. The student explains his/her major subject in some detail to the group, who probably have only a general idea of it. Subjects include mathematics, social studies, music, art, botany, home economics, etc. This task is more on the order of an informal oral presentation. The purpose is to see how well the student can handle professional interaction in English. Expectations are that students will feel most at ease (or least uncomfortable) talking about a subject that they know well, care about, and have a lot to contribute.

Communicating one's subject is central to the teaching profession. A teacher must be able to draw from a range of appropriate words, phrases, and approaches in English to explain the subject to others, especially to the pupils. The ability to summarise, condense, and simplify information is essential. The language needed is professional in tone.

Phase III: Argumentative discourse. Both students respond to a contentious statement on a current topic. Professionals are educated citizens who must be aware of issues facing South African society in transformation. They must be able to take a stand and support it, which involves persuasive strategies as well as language that tends to be more abstract than concrete. A few topics are given in the testing situation, and the students can decide which to discuss. Examples include the content of proposed censorship laws, re-organization of schools, and abortion laws. This phase can bring out emotion-laden language.

SCORING OF THE TOPISA

Evaluators use a Rating Scale independently to rate each student on each phase. Afterwards they confer to consolidate ratings for a single score per student. The scale is as follows:

45% - Below standard for teaching in English in South Africa

50% - Able to manage and give instruction in English at a very basic level ("little e")

55% - Able to manage and give instruction in English at a modest level ("little e")

60% - Able to manage and give instruction in English at a competent level ("big E")

65% - Able to function in English at near mother-tongue level of competence.

The scoring scale is based on percents in the marking system in general use in South Africa, in which 50% signifies the basic passing mark, 55% a better pass, 60% a good, decent pass, and 65% signifies a fine pass. In other words, under 50% indicates insufficient English language proficiency to teach anything through the medium of English. A 65% signifies sufficient proficiency to teach anything through the medium of English. Marks lower than 50% or higher than 65% can be assigned, but in practice are used only rarely.

The 15 points in between are given several meanings. The basic pass of 50% gives the student the "little e". The 55% gives the student still the "e", along with some encouragement that he/she is better than basic but not good enough for all requirements, and he/she needs to work at English and perhaps try again in six months. The 60% gives the student the "big E", or final approval of English L2 proficiency.

In practice, this scale offers three main differentiations: no pass, partial pass (e) and full pass (E). The (e) and (E) are further subdivided into two levels each, a lower and an upper level. It seems there are five intervals of assumed unequal size.

These ratings are awarded by consensus of the two interviewers/evaluators.

THEORETICAL DESCRIPTION

Nature of the test. The TOPISA continues the oral language assessment approach developed in the 1950's at the Foreign Service Institute (FSI) in Washington, DC. The FSI test has become the Interagency Language Roundtable (IRL) test, developed as a multi-language assessment device in the USA diplomatic service for professionals who must learn to speak and read a country's language before they are sent there to work. The IRL test has spawned many other efforts, notably one for high-school foreign language learners which has also become quite influential, the American Council on the Teaching of Foreign Languages (ACTFL) test. The ACTFL test has distinguished several levels of skill at beginning stages of learning a language.

Such tests elicit a sample, hopefully a representative one, of the speaking and listening skills that a learner has acquired. This sample is rated against certain requirements by an evaluator trained to make this rating. For the TOPISA, the general requirement is sufficient L2 skill to teach a primary or secondary school class in the medium of English. What "sufficient" means in this context is still a rather subjective matter.

Characteristics. The TOPISA is a direct test as opposed to semi-direct or indirect (Clark, 1979). The learner actually sits down and listens and speaks with other people and is assessed on that basis. By contrast, an indirect or semi-direct measure elicits speech by means of pictures, tape recordings, or other non-human means (Clark, 1979). The TOPISA is based on authentic human interaction with living human beings in a specific context.

The TOPISA is not a structured oral test in the sense that the items are the same to all interviewees with an accepted set of answers (Jones, 1985). It is more a testing procedure than a structured test. Jones (1985: 80) calls the interview test "the highest form of oral testing."

The TOPISA is flexible, in that there are no set questions or topics that must be dealt with, no specific desired answers that can be right or wrong. It is open-ended regarding topic, style, and tone. It is a global rather than a discrete-item measure. However, it is structured to a minimum degree in that three kinds of discourse are built into it. It is adjustable in that evaluators can reduce its difficulty to assure successful interaction, as the limits of the students so indicate.

Its range is deliberately limited to only upper-level skills. Students are called to take it only when they have completed three years of university work and have passed the written certification test at the level of the "big E". Thus the TOPISA operates at the upper end of oral proficiency. One could hazard a guess that students who may score at least 2+, or about midway, on the IRL scale an up would be called for this test.

It is deliberately limited in domain as well. It deals only with that subset of English that is likely to be used by a teacher in the course of his/her work, as described earlier. This limitation of domain may make the TOPISA more amenable to scalability, or ordering of levels along a continuum according to, for example, function, content, or accuracy.

Description in terms of test theory has assisted in providing some tools for a more critical look at the TOPISA, given the need for such a measure to be extended to other institutions charged with certifying teachers in second-language proficiency.

FURTHER RESEARCH AND DEVELOPMENT OF THE TOPISA

Having conceptualised the TOPISA, implemented it, and verified its utility, it seems necessary to come to a more scientific examination of it. This exploration should lead also to its improvement.

1 SKILL DESCRIPTIONS

Skills and functions need to be more clearly identified so that they are seen by all stakeholders to represent fairly and comprehensively the range of communicative tasks in English required of teachers.

The criteria for the five levels of the TOPISA at this point are more implicit than explicit. They have developed from the experience and education of many English Education faculty members who have had decades of extensive experience with all aspects of English language education in South Africa, from research to classroom instruction to teacher preparation to country-wide examination preparation and control. They know what oral English skills teachers need. They are capable of identifying the level of L2 proficiency of pre-service teachers with great ease and accuracy. The problem comes in when it becomes necessary to communicate these standards to others, such as to faculty from outside the department or to students including pre-service teachers. Then it would be extremely useful to be able to express the standards in written form.

This is the next step in developing the TOPISA. If the TOPISA is to be extended, professionals from other institutions as well as the Department of Education could be invited to help define the levels in written form.

As recommended by Bachman and Clark (1987) cited in Clark and Lett (1988), the first step in a systematic investigation of testing issues is to develop a prototypical model of communicative language ability, for teachers in this context. This would be a rather comprehensive model. Then performance tests covering components in this model could be developed and subjected to tests of validity and reliability. The various performance tests could be synthesized into one measure.

2 VALIDITY

Henning (1987) in his overview of language testing in general offers several kinds of validity considered applicable to language tests. It seems that the TOPISA by its nature, characteristics, and application, would not have some kinds of validity concerns at issue: face validity (it has intuitive approval by its users), response validity (examinees are generally cooperative), and concurrent validity (similar measures of known validity do not exist).

Other types of validity may need to be demonstrated to some extent for the TOPISA. Content validity (Does it represent the full range of L2 skills needed by teachers?) seems at least partially established, in-that there can be no doubt that the three discourse types tested are needed by teachers. This observation has the general agreement of department members. However, is the test comprehensive enough? Are there essential skills left unexamined? How does the TOPISA take into account variations within South African English(es)? To what extent can/should acceptability levels be negotiated?

Predictive validity (correlation with some measure of success in the field) is of great interest, in that the TOPISA must identify persons who can function in the future in their L2. To date, no one who has obtained the "big E" through the Department has been

BEST COPY AVAILABLE

known to lack sufficient L2 proficiency to handle his or her teaching situation. A further question arises: Is the TOPISA perhaps too strict?

The last major type of validity, according to Henning (1987), is construct validity, which itself has several possible ways of being established. Using a general approach, one first establishes one or more constructs that derive from formal theory. Predictions are made from that theory and tested. If the predictions do indeed occur, the hypothesis is supported and the construct can be said to be validated. If the constructs which underly the TOPISA are identified, tested, and supported, the TOPISA can be said to have construct validity. Perhaps the three kinds of discourse can lead to construct formulation.

Validity asks what is in this test. Is it defined by an ad hoc process rather than an a priori, principled, and generalizable approach (Clark & Lett, 1988)?

The validity of any measure must be shown in a systematic and scientific way, even while users have confidence in the validity of the TOPISA as it is currently implemented.

3 RELIABILITY

The reliability of the TOPISA (its consistency of measurement) must be established. Will it yield the same results, no matter who the interviewers/evaluators are? No matter which specific topics come up for discussion? No matter where it is given? Are there specific circumstances which it requires, or which would nullify it?

One of the threats of reliability listed by Henning (1987) is fluctuations in the learner (when re-tested, for example, or when ill or fatigued). Another is fluctuations in scoring, from within a given rater (intra-rater variance) or between raters (inter-rater variance). A third is fluctuations in the environment. Of these three, scoring variations seem to present an area for attention. Detailed rating schedules, re-evaluations, and rater training may become imperative if the TOPISA is extended. One check on reliability is the presence of two raters, who give a joint assessment.

Reliability can be studied with regard to length, difficulty, and boundary effects, as well as discriminability, speededness, and homogeneity.

Other aspects of reliability seem more applicable to discrete item tests and will not be discussed here.

ADDITIONAL QUESTIONS

1. What exactly contributes to the proficiency ratings? Fluency? Accuracy? Effort? Confidence? Ethnic differences? Social status? Attractiveness? Prior acquaintance? Gender differences? These have all been shown in other contexts to be influential on evaluators' judgements.

2. Can the TOPISA be adapted (perhaps as an alternate form) from a direct to a semi-direct measure, using taped interviews which can then be rated at a later time and other place? There are many advantages, should this be possible. The Guam Educators' Test of English Proficiency (GETEP) could serve as a model here (Stansfield, Karl & Kenyon, 1990).

3. Can the TOPISA procedure itself be adapted to assess teacher preparedness in the other 10 official languages in South Africa? This would resemble the ILR test content. If so, validity and reliability for each target group would need to be established.

REFERENCES

Alderson, J.C., Krahnke, K.J. & Stansfield, C.W. eds. 1987. REVIEWS OF ENGLISH LANGUAGE PROFICIENCY TESTS. Washington, D.C.: Teachers of English to Speakers of Other Languages (TESOL).

Clark, J.L.D. 1979. Direct vs. semi-direct tests of speaking ability. In E.J. Briere & F.B. Hinofotis (eds), CONCEPTS IN LANGUAGE TESTING: SOME RECENT STUDIES, 35-49. Washington, DC: TESOL.

Clark, J.L.D. & Lett, J. 1988. A research agenda. In Lowe, P. & Stansfield, C.W.

Henning, G. 1987. A guide to language testing: development, evaluation, research. New York: Newbury House.

Jones, R.L. 1985. Some basic considerations in testing oral proficiency. In NEW DIRECTIONS IN LANGUAGE TESTING. Oxford: Pergamon Press.

Lowe, P. & Stansfield, C.W. 1988. SECOND LANGUAGE PROFICIENCY ASSESSMENT: CURRENT ISSUES. Englewood Cliffs, New Jersey, USA: Prentice Hall.

Stansfield, C.W. 1993. An approach to performance testing of oral language proficiency for bilingual education teacher certification. In PROCEEDINGS OF THE THIRD NATIONAL RESEARCH SYMPOSIUM ON LIMITED ENGLISH PROFICIENT STUDENT ISSUES: FOCUS ON MIDDLE AND HIGH SCHOOL ISSUES. Washington, D.C.: USA Department of Education, 187-215.

Stansfield, C.W. Karl, J. & Kenyon, D.M. 1990. THE GUAM EDUCATORS' TEST OF ENGLISH PROFICIENCY (GETEP). Washington, D.C.: Center for Applied Linguistics.